

Die zweite Diskussionswelle über die Auswertungsprobleme eines 3-Wellen-Panels kategorischer Daten: einige Anmerkungen zu der Entgegnung von Jagodzinski

Langeheine, Rolf

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Langeheine, R. (1987). Die zweite Diskussionswelle über die Auswertungsprobleme eines 3-Wellen-Panels kategorischer Daten: einige Anmerkungen zu der Entgegnung von Jagodzinski. *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, 21, 70-76. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-205270>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.



Die zweite Diskussionswelle über die Auswertungsprobleme
eines 3-Wellen-Panels kategorischer Daten: Einige Anmerkungen
zu der Entgegnung von Jagodzinski

von Rolf Langeheine

JAGODZINSKI's (1986) Arbeit in der ZA-Information 19 hat zu einer Diskussion zwischen mir (LANGEHEINE 1987) und JAGODZINSKI (1987) geführt, die in der ZA-Information 20 nachzulesen ist. Solche Diskussionen, Debatten oder Streitgespräche sind - gerade in wissenschaftlichen Foren - deshalb besonders interessant, weil sie die Autoren zwingen, Dinge auf den Punkt zu bringen. Häufig entsteht damit die Chance, daß alle - Leser wie Autoren - mehr bei der Sache lernen (jedenfalls kann ich es so in diesem Zusammenhang für mich sagen). Ich möchte im folgenden zunächst zeigen, daß JAGODZINSKI's und meine Ergebnisse in der Tat dramatisch differieren, und dann auf die beiden Thesen von JAGODZINSKI (1987) eingehen.

Zum Punkt Reliabilität/Stabilität:

Bei fast allen sozialwissenschaftlichen Datenerhebungen müssen wir davon ausgehen, daß Meßfehler eine Rolle spielen. Dies gilt auch für kategoriale Daten (vgl. z.B. BYE & SCHECHTER, 1986; SCHWARTZ, 1985; VAN DE POL & DE LEEUW, 1986). Im Fall von kategorialen Paneldaten bieten Latent-Class-Modelle (LCM'e) nun die Möglichkeit, die beiden Aspekte Reliabilität und Stabilität zu trennen (ich betone dies, da es in meinem vorangegangenen Papier nicht deutlich genug geworden ist). Unter Modell M4 z.B. läßt sich die bedingte Wahrscheinlichkeit von .876 als Reliabilität für die latente Klasse der Materialisten zum Zeitpunkt t1 interpretieren. Idealerweise (d.h. bei Annahme perfekter Reliabilität) wäre die 3x3 Matrix (Kategorien x Klassen) zum Zeitpunkt t1 eine Einheitsmatrix (wie z.B. für die Klassen 1-3 unter Modell M2). Entscheidet man sich für diese Interpretationsmöglichkeit von LCM'en, so lassen sich die Parameter eines Modells dazu verwenden, detaillierte Fehlerraten (Zeitpunkt je Klasse) oder generellere Fehlerraten (Zeitpunkte bzw. Klassen) zu schätzen. Diese Fehlerraten differieren unter den Modellen M3 und M4 in der Tat erheblich (M3: .604, .517, .517; M4: .230, .170, .146 - jeweils für t1-t3. Der Grund ist schlicht der, daß sich die beiden Modelle in den jeweils getroffenen Annahmen stark unterscheiden (damit ist noch nichts über die Adäquatheit dieser Modelle gesagt). Der Anteil der Zeitstabilen ergibt sich dagegen je Klasse aus dem mit dem jeweiligen Klassenanteil gewichteten



Produkt der Reliabilitäten. Unter Modell M4 sind dies insgesamt 55.4%, unter der LCM-Version von M3 erwartungsgemäß erheblich weniger, nämlich nur 39.5%.

Zu JAGODZINSKI's These 1:

Es ist unbestritten, daß die X^2 -Approximation für kleines N und dünn besetzte Zellen in Zweifel gezogen werden kann. Trotzdem muß man deshalb nicht, wie wir noch sehen werden, statistische Tests gleich ganz über Bord werfen. Das Problem hat eine Reihe von Statistikern veranlaßt, in Monte Carlo Studien die X^2 -Approximation für die Teststatistiken L^2 (Likelihood Ratio Chi-Quadrat) und X^2 (Pearson Chi-Quadrat) zu untersuchen. Das Ergebnis dieser Studien ist, daß X^2 der X^2 -Approximation häufig (aber eben nicht immer) besser folgt als L^2 . Das Problem bei der Übertragung dieser Resultate auf den konkreten Fall besteht allerdings darin, daß für das gerade zur Diskussion stehende Modell meistens keine entsprechenden Monte Carlo Ergebnisse vorliegen. Dies gilt für alle von uns betrachteten "Modelle. Streng genommen kann man sich also auf den Standpunkt stellen, daß die Ergebnisse solcher Simulationsstudien in der hier vorliegenden Situation absolut nicht weiterhelfen, da man über den konkreten Fall nichts weiß. Man kann allerdings einen weichen Standpunkt einnehmen und die generellen Trends dieser Studien nutzen, in der Annahme, sie würden auch für den je konkreten Fall gelten. Der Grund für das häufige Fehlen von Monte Carlo Studien ist der, daß zur Abdeckung aller möglichen Fälle in Monte Carlo Studien derart viele Faktoren variiert werden müßten, daß ein solches Unternehmen von Anbeginn zum Scheitern verurteilt wäre. Die Zahl dieser Faktoren, die in ausgewählten Kombinationen für loglineare Modelle untersucht wurden, erhöht sich bei LCM'en zusätzlich. Es ist daher nicht verwunderlich, wenn es entsprechende Studien für LCM'e nicht gibt. Die einzige mir bekannte Monte Carlo Studie zu LCM'en (POULSEN, 1982) galt der Güte der Parameterschätzungen. Und die sind - auch für kleines N - dann ohne Bias, wenn sich die Klassen in ihren Profilen (bedingte Wahrscheinlichkeiten) gut trennen lassen. Da dieses für alle von mir untersuchten Modelle zutrifft, gibt es also ein gewisses Indiz dafür, daß wir den erwarteten Häufigkeiten (die ja auf den Modellparametern beruhen) trauen können. Das ist wenigstens etwas. Es sagt aber noch nichts über die Angemessenheit der Chi-Quadrat Statistiken. Die Frage ist nun, was man macht, wenn man trotz kleiner erwarteter



Zellhäufigkeiten auf statistische Tests nicht verzichten möchte. Die Antwort (für mich) ist, daß es zwar keinen Königsweg, aber doch eine intersubjektiv überprüfbare Strategie gibt. Ich habe mich dabei an die Regeln von FIENBERG (1979) gehalten, die das Fazit der von ihm herangezogenen Monte Carlo Studien sind (1. die minimale erwartete Häufigkeit sollte 1 betragen; 2. N sollte wenigstens das 4- bis 5-fache der Anzahl der Zellen der Tabelle betragen), und habe Zellen mit Erwartungswert kleiner 1 zusammengefaßt. Es sei angemerkt, daß es für derartige Gruppierungen (Zusammenfassungen) verschiedene Strategien gibt. So kann man z.B. alle Zellen zusammenfassen, die das festgesetzte Minimum unterschreiten. Oder man kann so lange gruppieren, bis das Minimum erreicht ist und dann neu beginnen (und sich dabei noch überlegen, ob benachbarte Zellen gruppiert werden sollen oder ob dies nach Zufall geschehen soll). Im allgemeinen scheint dies aber nicht zu unterschiedlichen Konsequenzen zu führen. Hingegen scheint die Gruppierung mit zunehmendem Minimum (1, 3, 5) nach den Ergebnissen von ROST (1988) einen antikonservativen Effekt auf die Chi-Quadrat-Statistiken zu haben (d.h. ein Modell wird schneller abgelehnt). Dies deckt sich mit FIENBERGSs (1979) Fazit.

Da nun die korrigierten L^2 -Werte für keines der von mir untersuchten Modelle zu einer anderen Schlußfolgerung führten als die unkorrigierten Werte, fühlte ich mich relativ auf der sicheren Seite. Ich habe allerdings bislang (wegen der angenehmen Eigenschaft der Partitionierbarkeit von L^2) nicht von der Möglichkeit Gebrauch gemacht, auch das Pearson X^2 zur Prüfung der Modellgültigkeit heranzuziehen, was von einigen Autoren generell empfohlen wird. In Tabelle 1 finden sich daher unkorrigierte und korrigierte L^2 - und X^2 -Werte für alle von mir untersuchten Modelle. Diese Tabelle zeigt nun, daß alle Test-Statistiken auf dem 5%-Niveau zu gleichen Schlußfolgerungen führen - mit einer Ausnahme: das Modell M3 (also JAGODZINSKI's Modell) wird nach L^2 verworfen, nach X^2 hingegen akzeptiert. Da beide Statistiken nicht zugleich richtig sein können, läßt sich über die Nicht-/Angemessenheit von M3 nur anhand neuer Daten entscheiden.

Dennoch sind wir damit noch nicht ganz am Ende der Reise. Trotz der Zweifel, die bei der vorliegenden Datenkonstellation an den Statistiken für einzelne Modelltests angebracht sind, ermöglicht die L^2 -Differenz zweier hierarchisch geschalteter Modelle einen validen Test der H_0 , daß beide

Modelle die Daten gleich gut anpassen (für Details s. z.B. HABERMAN,



Tabelle 1: L^2 und X^2 Teststatistiken für die in Langeheine (1987) betrachteten Modelle

Modell	L^2		X^2		df	
	unkorr.	korr.	unkorr.	korr.	unkorr.	korr.
M2	42.45	38.21	36.09	33.20	23	18
M3	31.81	27.52	27.20	23.58	19	14
"Mover-Stayer"	38.29	34.30	37.90	35.07	17	12
Third Force	30.28	26.12	31.14	27.32	18	13
M4	13.81	7.04	14.14	4.41	12	3
M5	38.82	29.92	40.54	24.95	18	12
M6	21.24	8.42	26.42	8.20	20	9
M7	25.63	12.99	27.25	12.10	23	12

1978, S. 325 ff.). Somit ist zumindest eindeutig belegbar, daß die Anpassung unter M4 signifikant besser ist als unter M3 (ebenso paßt M3 signifikant besser als M2). Da M4 3 und M3 4 Klassen hat, mag es zunächst so erscheinen, als ließe sich das restriktivere Modell M3 nicht durch Restriktionen aus M4 ableiten. Man kann jedoch ein Modell M4' spezifizieren, das neben den Klassen 1-3 wie unter M4 zusätzlich eine Zufallsklasse wie die Klasse 4 von M3 enthält. Dann ist M3 ein Submodell von M4'. Aufgrund der Datenkonstellation zeigt sich nun in diesem speziellen Fall, daß M4 identisch ist mit M4', da die Zufallsklasse von M4' keine Fälle enthält. Dies muß natürlich nicht für neue Daten gelten. Das von JAGODZINSKI (1987) unter Punkt 3 als angemessen erachtete Modell (M4' also) ist somit grundsätzlich sinnvoll.

Als Fazit ergibt sich also, daß die Verwendung von L^2 zur Beurteilung der Anpassungsgüte einzelner Modelle zweifelhaft sein mag, daß L^2 aber sehr wohl als inferenzstatistisches Maß zur Diskriminierung der Modelle M2 - M4 geeignet ist. Aufgrund der Meßfehlerproblematik erscheint es mir im übrigen auch wenig sinnvoll, Meßfehler für die Klassen 1-3 nur zum Zeitpunkt t1 zuzulassen, diese aber für die weiteren Zeitpunkte in eine separate Klasse zu zwingen, in der der pure Zufall regiert. JAGODZINSKI (1987) sagt unter Punkt 3 selbst, daß dafür allein methodisch-technische Gründe verantwortlich waren.

Zu JAGODZINSKI's These 2:

Geht man von den Daten, d.h. der vorliegenden 3x3x3 Kreuztabelle aus - und das mußte JAGODZINSKI (1986) auch bei den Berechnungen für seine Modelle M1 bis M3 -, so werden (mit Ausnahme saturierter Modelle und explizit über Restriktionen perfekt angepaßter Zellen wie z.B. unter M2 und M3) immer bestimmte Zellen über- bzw. unterschätzt. JAGODZINSKI's (1987) Beobachtung, daß unter M4 39.3 zeitstabile Mischtypen geschätzt werden, während es nur 38 beobachtete Fälle für dieses Antwortmuster gibt, kann daher M4 nicht in Frage stellen. Um M4 aber abzusichern, möchte ich anfügen: Erstens wurden für M4 mehrere MLLSA-Läufe mit verschiedenen Startwerten, aber gleichen Ergebnissen durchgeführt. Es kann also ausgeschlossen werden, daß der iterative Prozeß in einem lokalen Maximum endete. Zweitens habe ich M4 jetzt mit einem anderen Programm (LAT; HABERMAN, 1979) angepaßt. Das Ergebnis ist identisch mit dem aus MLLSA. Der Gedanke hinter diesem Check war, daß der Identifikationstest in MLLSA zuweilen nicht greift, während man sich auf diesen Test in LAT verlassen kann. M4 ist also identifiziert (und die Ergebnisse ändern sich auch nicht, wenn man ein wesentlich schärferes Konvergenzkriterium fordert).

Wenn JAGODZINSKI nun meint, daß M4 fehlspezifiziert ist, weil er zeigen kann, daß dieses Modell die Ausgangsdaten (desaggregierte Daten) schlechter reproduziert als das Modell M3, dann kann dies auf ein Aggregationsproblem hinweisen. Die desaggregierten Daten bestehen aus einer 12³-Tabelle (vgl. JAGODZINSKI 1986, Tabelle 1B und S. 43), die zu einer 3³-Tabelle aggregiert wurden - und nur auf diese 3³-Tabelle beziehen sich die von mir untersuchten Modelle. Man kann jedoch - was wir jetzt getan haben - ein niedrigeres Aggregationsniveau wählen und die 12 Antwortsequenzen (vgl. JAGODZINSKI 1986, Tabelle 1B) wie folgt zu 6 Kategorien zusammenfassen: (2)+(7), (6)+(11), (1)+(4), (3)+(10), (5)+(8), (9)+(12). Dann resultiert eine 6³-Tabelle mit Materialisten, Postmaterialisten und 4 Mischtypen. Analysiert man diese Tabelle mit den Modellen M3" und M4", die jeweils analog zu den Modellen M3 bzw. M4 gebaut sind (jetzt aber 7 bzw. 6 latente Klassen enthalten), so zeigt sich hinsichtlich Zeitstabilität das in Tabelle 2 wiedergegebene Ergebnis.

Interessant an diesem Ergebnis ist, daß sich beide Modelle hinsichtlich der Schätzung der zeitstabilen Gruppen praktisch nicht unterscheiden, obwohl



Tabelle 2: Stabilität für Materiallisten, Postmaterialisten und 4 Mischtypen für die Modelle M3" und M4"

	M3"		M4"	
	%	N	%	N
Materialisten	15.60	23.72	14.96	22.74
Postmaterialisten	14.29	21.72	13.69	20.80
Mischtyp 1	.47	.72	1.29	1.96
Mischtyp 2	8.37	12.72	9.48	14.41
Mischtyp 3	2.44	3.72	2.25	3.42
Mischtyp 4	1.79	2.72	2.08	3.15
Summe	42.96	65.32	43.75	66.48

Mischtyp 1 = Antwortsequenz (1)+(4)
Mischtyp 2 = Antwortsequenz (3)+(10)
Mischtyp 3 = Antwortsequenz (5)+(8)
Mischtyp 4 = Antwortsequenz (9)+(12)

es eine Differenz in der Modellanpassung gibt ($L^2_{M3''} = 171.82$, $df = 194$; $L^2_{M4''} = 78.87$, $df = 166$ - in beiden Fällen ohne Korrektur für erwartete Häufigkeiten kleiner 1). Der Schluß aus diesem Ergebnis kann nur sein, daß durch die Aggregation auf die 3³-Ebene relevante Information verlorengegangen ist, so daß sich die Analyse auf dieser Ebene - im Nachhinein betrachtet - verbietet.

Zum Schluß: Es mag plausible Gründe dafür geben, im vorliegenden Kontext auf ein Modell zu bauen, das (wie M2 in extremer Weise) Meinungslosigkeit von Befragten annimmt. Die mögen sogar theoretisch eher auf der Hand liegen als die Annahme, daß beobachtete Fluktuationen durch Antwortunsicherheit (also Meßfehler wie in M4) begründet sind. Auf der anderen Seite können wir solche Meßfehler nicht ausschließen. Ich sehe einen Vorteil der inzwischen vorhandenen statistischen Ressourcen darin, daß wir solche konkurrierenden Hypothesen gegeneinander testen können.

Literaturverzeichnis

- BYE, B.V. & SCHECHTER, E.S.: A latent Markov model approach to the estimation of response errors in multivariate panel data. *Journal of the American Statistical Association*, 1986, 81, 375-380.
- FIENBERG, S.E.: The use of chi-square statistics for categorical data problems. *Journal of the Royal Statistical Society, Series B*, 1979, 41, 54-64.
- HABERMAN, S.J.: Analysis of qualitative data. Vol 1. Introductory topics. New York: Academic Press, 1978.
- HABERMAN, S.J.: Analysis of qualitative data. Vol. 2. New developments. New York: Academic Press, 1979.
- JAGODZINSKI, W.: Black & White statt LISREL? Wie groß ist der Anteil von "Zufallsantworten" beim Postmaterialismusindex? *ZA-Information* 19, 1986, 30-51.
- JAGODZINSKI, W.: Über einige Anwendungs- und Interpretationsprobleme "anspruchsvoller" Schätzverfahren. Entgegnung auf den Beitrag von LANGEHEINE. *ZA-Information* 20, 1987, 56-63.
- LANGEHEINE, R.: Black & White, anfängliche Antwortunsicherheit, Mover-Stayer, Third Force oder was? Ein paar weitere Überlegungen zu JAGODZINSKI's Analyse des Postmaterialismus-Panels. *ZA-Information* 20, 1987, 44-55.
- POULSEN, C.S.: Latent structure analysis with choice modeling applications. (Doctoral Dissertation.) Aarhus: Aarhus School of Business Administration and Economics, 1982.
- ROST, J.: Test theory with qualitative and quantitative latent variables. Erscheint in R. LANGEHEINE & J. ROST (Hrsg.): *Latent trait and latent class models*. New York: Plenum, 1988.
- SCHWARTZ, J.E.: The neglected problem of measurement error in categorical data. *Sociological Methods & Research*, 1985, 13, 435-466.
- VAN DE POL, F. & DE LEEUW, J.: A latent Markov model to correct for measurement error. *Sociological Methods & Research*, 1986, 15, 118-141.

Dr. Rolf Langeheine
IPN an der Universität Kiel
Olshausenstr. 62
2300 Kiel 1